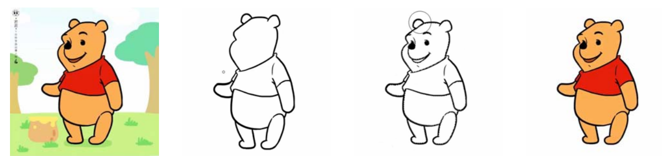
# Salient Contour-Aware Based Twice Learning Strategy for Saliency Detection

Chunbiao Zhu[1], Wei Yan[1], Shan Liu[2], Thomas Li[3], Ge Li[1,*]
[1]Peking University Shenzhen Graduate School
[2]Tencent America
[3]China and Advanced Institute of Information Technology, Peking University
[*]Corresponding author:geli@ece.pku.edu.cn

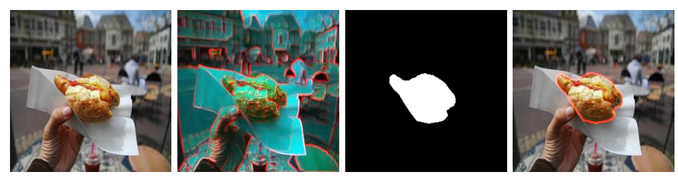
## Abstract

*Fully convolutional neural networks (FCNs) have shown outstanding performance in many computer vision tasks including salient object detection. However, most deep learning-based saliency detection models are too complicated. They cause difficulties in training. Additionally, the performance of those overly complex deep learning models is limited, and the price performance ratio of those complex models is very low. To address the problems of existing deep-learning-based methods, we introduce a new research field called saliency contour detection and design a new dataset for saliency contour detection. Inspired by the human sketching process, we propose a novel contour-aware algorithm using FCNs with a twice learning strategy for saliency detection, which imitates and dissects the process of human cognition. Extensive experimental evaluations demonstrate the effectiveness of our proposed method against other outstanding methods.*



(a) Sample (b) Framework (c) Details (d) Colour

Figure 1. Painting cognitive process.



(a) Image (b) Edge (c) Saliency (d) Contour

Figure 2. The difference between image edge and image salient contour.

## 1. Introduction

The human visual system possesses a focusing ability, which can rapidly search and screen out the salient object in various scenes. This kind of process is called the attention mechanism. Today, in order to solve the urgent problem of efficiently extracting useful data from the complicated data on the internet, we should learn from the human visual attention mechanism. Saliency detection is a basic research topic in computer vision. It conforms to the human visual cognition process and can simulate the human visual system to process visual information. Saliency detection can be applied in many fields, such as object tracking[28], image compression[27] etc. A number of saliency detection models have been proposed in the past two decades[3]. Although much effort has been made, saliency detection is still a challenging problem.

Traditional saliency detection methods employed low-level features and heuristic priors which are neither robust enough to discover salient objects in complicated scenes, nor able to capture high-level semantic knowledge. With the development of deep neural networks (DNNs), the drawbacks of traditional methods have been partially overcome. They can capture high-level semantic information over hand-crafted features, and thus are more effective in locating salient regions, leading to more accurate results in complicated scenes.

However, most deep learning-based saliency detection models are too complicated, and cause difficulties in training. Training such models requires a large dataset and is time-consuming. At the same time, the performance of those overly complex deep learning models is limited, and the price performance ratio of those complex models is very low. It is the lack of understanding of the human brain cognitive process that causes those problems. Figuring out the simplest network model and the strategy that fits the human brain cognitive process as much as possible to solve the above problem is worthy of study.

The proposed method is based on the human brain cognitive process and is inspired by the human sketching process. As shown in Figure 1, when an artist wants to copy

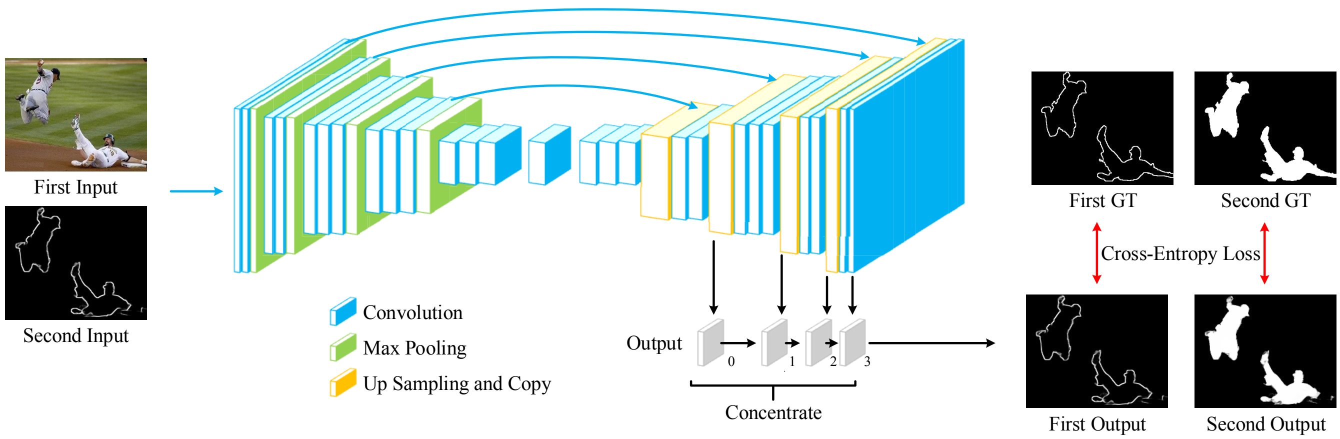



Figure 3. Overview of the main framework of our proposed approach.

an image, the operations to be performed in turn are: selecting salient objects, drawing a frame of salient objects, drawing internal details of salient objects, coloring salient objects, drawing scenes, and coloring. Inspired by the process of human copying, we believe that most of the traditional methods and neural network-based methods can not express the salient contour information of the image well (Figure 2), which often causes problems such as blurred edges and inaccurate contours. For this problem, this paper speculates that, first, the neural network does not follow the human brain cognitive process. The network does not understand the process of copying, and does not learn the contour information in advance, but rather directly detects the salient part of the image. Second, the direct learning of the saliency detection through the neural network is too complex, so the learning performance is not ideal. Therefore, this paper believes that the complex process of saliency detection should be simplified. The saliency detection should imitate the cognitive process of human brain. We should control the learning process of neural network, and gradually obtain the results of salient object detection.

In this paper, a novel contour-aware algorithm using FCN with twice learning strategy for saliency detection is proposed. We solve the problem of deep learning-based models being too complicated and difficult to train. This paper utilizes the characteristics of human brain cognition, and first proposes the research field of saliency contour detection. A new salient contour detection dataset is designed. At the same time, the proposed saliency detection algorithm based on the twice learning strategy is in line with the cognitive process of humans and simulates the painting process to detect salient objects. The proposed method is divided into two learning stages. The first learning stage is the saliency contour detection, and the second learning stage is the saliency filling process. In addition, these two learning phases are based on a simple and effective convolutional network designed by us. Supplemented by the twice learning strategy, the detection model is simple and effective. Experimental results also show that our proposed method achieves competitive performance.

## 2. Related work

Conventional methods employed hand-crafted visual features. Jiang et al. [11] formulate saliency detection via an absorbing Markov chain on an image graph model. Yang et al. [21] rank the similarity of the image regions with foreground cues or background cues via graph-based manifold ranking. More comprehensive analysis of these hand-crafted feature based methods can be found in [2]. Since the traditional methods are not robust in complex scenes, neither are capable of capturing semantic objects. Deep learning methods are used to overcome these problems. Much research efforts have been made to develop various deep architectures for useful features that characterize salient objects or regions. For instance, Zhu et al.[29] present a two-channeled perceiving residual pyramid networks for generating high-resolution and high-quality results for saliency detection. Li et al.[12] fine tune fully connected layers of multiple CNNs to predict the saliency degree of each superpixel. Li et al.[15] propose a FCN trained under the multi-task learning framework for saliency detection. Zhang et al.[23] present a generic framework to aggregate multi-level convolutional features for saliency detection. Deng et al[5] propose a recurrent residual refinement network equipped with residual refinement blocks for saliency detection. They learn the residual between the intermediate saliency prediction and the ground truth by alternatively leveraging the low-level integrated features and the high-level integrated features of a FCN. Although the proposed method is also based on deep learning, the main difference between ours and these methods is that they learn a model that directly detect the salient object, while our method is divided into the saliency contour detection and the saliency filling process.

## 3. Proposed approach

This section introduces our proposed method. Since our twice learning strategy is derived from the painting cognitive process, we first introduce the painting process. Then, we demonstrate the definition of saliency contour detection and the production of the saliency contour detection dataset. Finally, the twice learning network model is introduced.

### 3.1. Cognitive process of painting

To better understand our proposed method, we first introduce the cognitive process of painting. The process of the visual information of the human eye is usually divided into three stages. First, the human eyes receive the visual information. The second stage is the analysis of the input information received by the human visual cortex. The third stage is the process of the visual cortex of the brain and other related neurocortex integrating to process the analyzed information. During the second stage, the human visual cortex will perceive and classify the received light, and will first distinguish the basic elements of the image, such as color, line, curvature, and angle. In view of this cognitive process, it can also be understood from the process of human sketching. As shown in Figure 1, if a painter wants to paint a copy of of a object, the painter will first visually perceive the whole scene of the object to be copied. Then, after the visually received input information is transmitted to the human brain for visual nerve processing, the artist's brain will recognize it and determine the copying object to be painted. This process is also the saliency detection process mentioned in this paper. Next, the painter begins to draw the desired object, and first draws a frame of the copied object. After the frame is drawn, the details and scenes are added, and then the entire image is colored. Finally, a complete painting process is finished. The algorithm of this paper is inspired by this process. Therefore, a saliency detection method using the twice learning strategy with salient contour perception is proposed, and an attempt is made to simulate human brain visual cognition, which makes the learning process understandable.

### 3.2. Saliency contour detection

The proposed algorithm is inspired by the visual cognition process. We decompose the saliency detection process into a saliency contour detection process and a saliency contour filling process, thus obtaining a complete saliency detection process. This section will introduce the saliency contour detection.

The main purpose of research in the field of computer vision is to enable machines to have the ability to understand a vision of the real world. Although there are many tasks in computer vision, such as object recognition, object classification, object tracking, 3D reconstruction, etc., which enable machines to act like humans, these tasks are too complex to enable the machine to understand the real object of the task, so machines just learn to imitate. Therefore, it is necessary to explore a new image structure that is conducive to visual perception, and transform the complex image into an intermediate rendering state by extracting the structural information, so that the intermediate state can provide powerful semantic information for high-level tasks.


Figure 4. New labels of the saliency contour detection dataset.

This paper proposes the concept of saliency contour detection for the first time. Saliency contour detection, as shown in Figure 2, is different from traditional edge detection. The saliency contour describes the contour of the salient object. Saliency contour detection is the process of detecting the external contour of a salient object in the image. By detecting the saliency contour information, it can play an auxiliary role for subsequent high-level computer vision tasks.

Since there is no research in this field, this paper also proposes a large-scale saliency contour detection dataset, which is modified and integrated according to the existing data set of saliency detection. The dataset acquisition process is shown in Figure 4. First, we choose 5 authoritative saliency detection datasets, which are MSRA10K[4], DUT-OMRON[21], DUTS-TE[20], ECSSD[18] and HKU-IS[13], containing 10,000 images, 5168 images, 5017 images, 1000 images, and 4447 images, respectively. Subsequently, the Canny detection operator[8] is used to detect the contour of the ground truth of those saliency detection datasets. Finally, we get the ground truth of the saliency contour detection dataset. In this paper, the saliency contour detection dataset is divided into two categories. The first category is used for training and validation, called PKUSCA-TR. The training set consists of the MSRA10K dataset with the newly marked saliency contour truth map, containing 10,000 images. The second category is used for testing, called PKUSCA-TE, which consists of DUT-OMRON, DUTS-TE, ECSSD, and HKU-IS datasets with newly labeled saliency contour truth maps, containing 15632 images.

### 3.3. Twice learning network model

#### 3.3.1 Master network architecture

The details of the proposed saliency detection network architecture are illustrated in Figure 3. Our master network is based on the encoder-decoder architecture. It consists of a contracting path and an expansive path. We use VGG [19] as the encoder part of the proposed model. Additionally, the copy-crop and multi-feature concatenation technique is used in the proposed model. We utilize hierarchical features in an effective way.

The details of our master network are as follows. Each of the convolution layers is followed by a Batch Normalization (BN) layer for improving the speed of convergence. To add non-linearity, we use the Rectified Linear Unit (ReLU) activation function. The convolution kernel size in our network is $3 \times 3$. By adding more low-level features from the early stage, the copy-crop technique is able to improve fine details of the saliency map during the up-sampling stage.

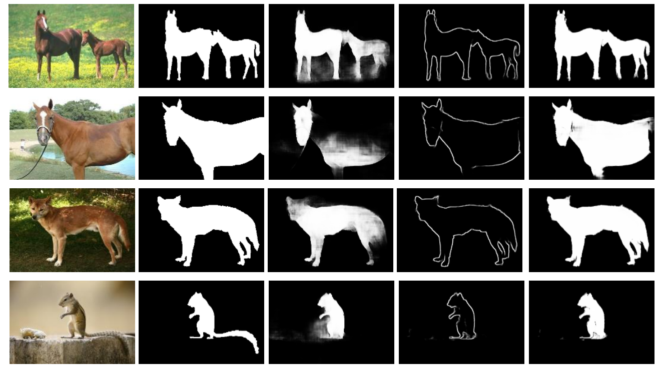
Based on loss-fusion pattern, multi-feature concatenation technique is used here for effectively combining both low-level and high-level features for loss fusion and accurate detection results. Those features in different blocks in decoder part go through one convolution kernel with $3 \times 3$ size and linear activation function, which produces pyramid outputs. They are concatenated to the final convolutional layer and the kernel size is $3 \times 3$. The sigmoid activation function is used in this layer. Finally, the pixel-wise binary cross entropy between predict saliency map $S$ and the ground truth saliency mask $G$ is computed by:

$$loss = \sum_{i=0}^{W} \sum_{j=0}^{H} \frac{(1 - G_{ij}) \cdot log(1 - S_{ij}) - G_{ij} \cdot log(S_{ij})}{W \times H}, \tag{1}$$

where $i, j$ are the pixel location in an image.

#### 3.3.2 Twice learning strategy

The proposed twice learning strategy is based on the saliency contour detection. As shown in Figure 3, for the first learning process, the input of the main framework is the original color image, and the output is the result of the saliency contour detection. This learning process can be seen as the salient contour learning process. For the second learning process, the input of the main framework is the ground truth of the saliency contour detection dataset mentioned in the section 3.2, and the output is the filled map of the saliency contour map. During the first training process, we adopt original color images and their corresponding saliency contour maps in PKUSCA-TR dataset mentioned in the saliency contour detection section. During the second training process, we adopt saliency contour maps and their corresponding saliency maps in PKUSCA-TR. The input information and the salient object of images used in the first training process and the second training process are the same. Only the ground truth of the output is different. During the inference process, the two trained networks are cascaded together to finish the whole saliency detection process.



(a) Image (b) GT (c) Our_SCA (d) Contour (e) Our_SCAT

Figure 5. The subjective results of the ablation study.

## 4. Experimental results

### 4.1. Datasets

For the training, we utilize the MSRA10K dataset[4]. It includes 10000 images with high quality pixel-wise annotations. We use Canny detection operator[8] to detect the contour of saliency maps, which are regarded as ground truths of the saliency contour detection. For the performance evaluation, we adopt 4 public large scale saliency detection datasets as follows.

**DUT-MORON**[21]. This dataset has 5,168 high quality images. Images of this dataset have one or more salient objects and relatively complex backgrounds.

**DUTS-TE** [20]. This dataset contains 5,019 test images with high quality pixel-wise annotations.

**ECSSD**[18]. This dataset contains 1,000 natural images, which include many semantically meaningful and complex structures in their ground truth segmentation.

**HKU-IS**[13]. This dataset contains 4,447 images with high quality pixel-wise annotations. Images of this dataset are well chosen to include multiple disconnected salient objects or objects touching the image boundary.

### 4.2. Ablation study

The proposed algorithm is mainly composed of four parts, namely the basic convolution-deconvolution ar-

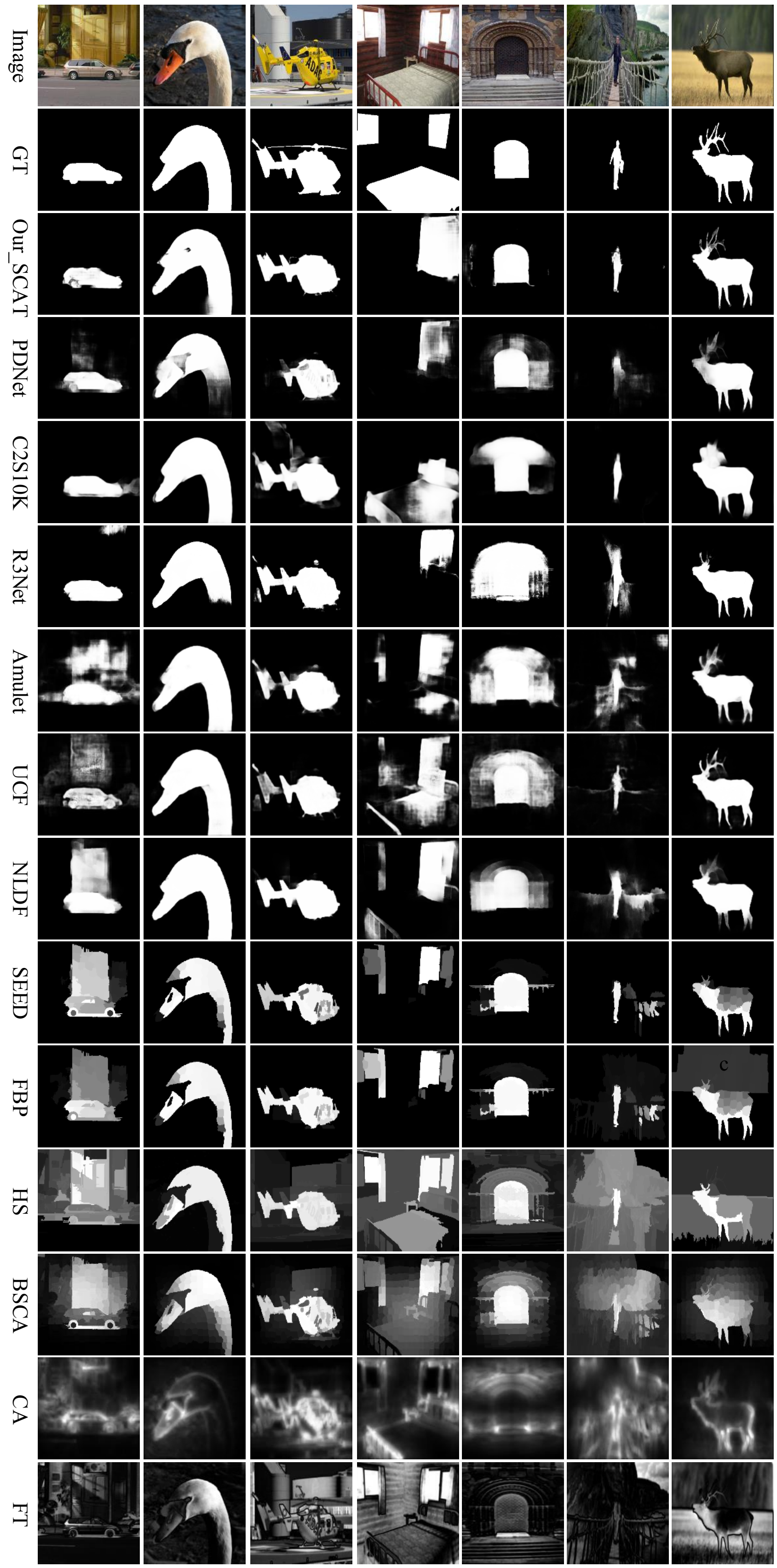



Figure 6. Comparison of subjective results.

chitecture, the Skip connection operation enhancement, the Concentrate cascade operation enhancement, and the twice learning enhancement. The basic convolution-deconvolution architecture, the skip connection operation enhancement, and the concentrate cascading operation enhancement parts are all existing neural network general models, which have been verified in an existing work[17]. Therefore, in this section, we only verify the effectiveness of the twice learning strategy.

Firstly, based on the saliency detection network model proposed in this paper, the main network is not modified, and only the input image and the output label and training processes are changed. Among them, the direct learning is defined as the Our_SCA process. In the direct learning training process, the input of the master network is the color image in the MSRA10K dataset[4], and the output is the result of the saliency detection. The ground truth in the MSRA10K is used to be compared with the output result. At the same time, the twice learning process is defined as the Our_SCAT process. It can be divided into the saliency contour detection process and the saliency filling process. The input of the saliency contour detection network is the color image in the PKUSCA-TR, and the output of it is the saliency contour map. The input of the saliency filling network is the saliency contour map in the PKUSCA-TR, and the output of it is the final saliency detection map. Then, we test the result of the Our_SCA and the Our_SCAT in DUT-OMRON[21], DUTS-TE[20], ECSSD[18] and HKU-IS[13]. We use four evaluation metrics, which are MAE, F-measure, E-measure[7] and S-measure[6]. The evaluation results are shown in Table 1.

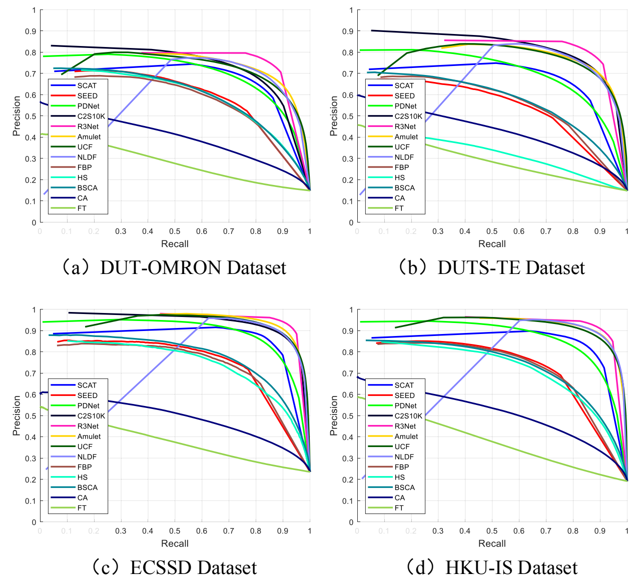
As seen from Table 1, the saliency detection method based on contour-aware twice learning strategy proposed in this paper is more effective than the direct learning. The performance of the twice learning strategy on all datasets is better than those of the direct learning at all evaluations. Results verify that the twice learning strategy based on contour perception makes a prominent contribution to the final saliency detection result. In addition, in order to better show that the twice learning is more effective than the direct learning for the final saliency test results, subjective assessment is also used to visually analyze the results of the direct learning and the twice learning. As shown in Figure 5, it can be seen that the twice learning process preserves the contour information better than the direct learning process, and the final saliency detection result is more accurate. It can be proven that the twice learning of the algorithm is a key step. Thus, the overall effectiveness of the algorithm is also verified.

## 4.3. Comparison results

We compare our algorithm with 12 other state-of-the-art methods including 6 deep learning based algorithms(PDNet[26], C2S10K[14], R3Net[5], Amulet[24], UCF[25], NLDF[16]) and 6 conventional algorithms(SEED[30], FBP[10], HS[18], BSCA[22], CA[9], FT[1]). To be fair, we use either the implementations with recommended parameter settings or the saliency maps provided by the authors. The results of the C2S10K on HKU-IS dataset is not presented because authors of this method can not provide their resulted saliency maps. As shown in table 2, our method outperforms the other methods across most of the datasets in terms of most evaluation metrics. From the PR curve(Figure 7), it can be seen that our method achieve competitive performance in all of the four datasets. Figure 6 provides a visual comparison of our approach with other state-of-the-art approaches. It can be seen that our proposed method has the finest detail as well as highlights the most correct salient region thanks to the salient contour-aware based twice learning strategy.

Table 1. Ablation study of the proposed twice learning strategy

| Method | Dataset | MAE | F-measure | E-measure | S-measure |
|---|---|---|---|---|---|
| Our_SCAT | DUT-OMRON | **0.084** | **0.650** | **0.806** | **0.767** |
| Our_SCA | | 0.095 | 0.612 | 0.770 | 0.757 |
| Our_SCAT | DUTS-TE | **0.094** | **0.637** | **0.796** | **0.752** |
| Our_SCA | | 0.107 | 0.596 | 0.756 | 0.742 |
| Our_SCAT | ECSSD | **0.077** | **0.828** | **0.885** | **0.842** |
| Our_SCA | | 0.093 | 0.805 | 0.864 | 0.831 |
| Our_SCAT | HKU-IS | **0.064** | **0.807** | **0.898** | **0.839** |
| Our_SCA | | 0.083 | 0.766 | 0.865 | 0.819 |



(a) DUT-OMRON Dataset (b) DUTS-TE Dataset
(c) ECSSD Dataset (d) HKU-IS Dataset

Figure 7. The PR curves on four datasets.

## 4.4. Failure case

Although the proposed method achieves competitive results on most comparison experiments, there is a limitation of the method. Figure 8 shows the failure case of the pro-

Table 2. Quantitative comparison of MAE, F-measure(F), E-measure(E) and S-measure(S) scores on large-scale RGB saliency detection datasets.

| Method | DUT-OMRON | | | | HKU-IS | | | | DUTS-TE | | | | ECSSD | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE | F | E | S | MAE | F | E | S | MAE | F | E | S | MAE | F | E | S |
| Ours | 0.084 | 0.650 | 0.806 | 0.767 | 0.064 | 0.807 | 0.898 | 0.839 | 0.094 | 0.637 | 0.796 | 0.752 | 0.077 | 0.828 | 0.855 | 0.842 |
| PD-Net | 0.095 | 0.612 | 0.770 | 0.757 | 0.083 | 0.766 | 0.865 | 0.819 | 0.107 | 0.596 | 0.756 | 0.742 | 0.093 | 0.805 | 0.864 | 0.831 |
| C2S10K | 0.079 | 0.664 | 0.817 | 0.780 | - | - | - | - | 0.066 | 0.710 | 0.841 | 0.817 | 0.059 | 0.853 | 0.906 | 0.882 |
| R3Net | **0.063** | **0.748** | **0.851** | **0.817** | **0.036** | **0.893** | **0.940** | **0.895** | **0.057** | **0.785** | **0.868** | **0.835** | **0.040** | **0.914** | **0.929** | **0.910** |
| Amulet | 0.098 | 0.647 | 0.779 | 0.781 | 0.051 | 0.841 | 0.912 | 0.886 | 0.085 | 0.678 | 0.794 | 0.804 | 0.059 | 0.868 | 0.901 | 0.894 |
| UCF | 0.120 | 0.621 | 0.765 | 0.760 | 0.062 | 0.823 | 0.902 | 0.875 | 0.112 | 0.631 | 0.762 | 0.782 | 0.069 | 0.844 | 0.892 | 0.883 |
| NLDF | 0.082 | 0.668 | 0.803 | 0.786 | 0.047 | 0.855 | 0.924 | 0.885 | 0.075 | 0.697 | 0.821 | 0.806 | 0.061 | 0.860 | 0.903 | 0.883 |
| SEED | 0.130 | 0.571 | 0.742 | 0.667 | 0.137 | 0.691 | 0.788 | 0.668 | 0.149 | 0.524 | 0.726 | 0.614 | 0.162 | 0.691 | 0.766 | 0.669 |
| FBP | 0.144 | 0.553 | 0.733 | 0.663 | 0.138 | 0.689 | 0.795 | 0.688 | 0.154 | 0.533 | 0.726 | 0.634 | 0.165 | 0.682 | 0.768 | 0.685 |
| HS | 0.227 | 0.519 | 0.705 | 0.633 | 0.215 | 0.638 | 0.761 | 0.674 | 0.284 | 0.296 | 0.631 | 0.498 | 0.228 | 0.634 | 0.724 | 0.685 |
| BSCA | 0.191 | 0.509 | 0.706 | 0.652 | 0.175 | 0.654 | 0.795 | 0.702 | 0.199 | 0.491 | 0.694 | 0.633 | 0.182 | 0.703 | 0.797 | 0.725 |
| CA | 0.255 | 0.355 | 0.655 | 0.536 | 0.274 | 0.460 | 0.696 | 0.572 | 0.259 | 0.373 | 0.662 | 0.545 | 0.309 | 0.430 | 0.650 | 0.543 |
| FT | 0.247 | 0.281 | 0.635 | 0.475 | 0.252 | 0.375 | 0.692 | 0.477 | 0.233 | 0.281 | 0.653 | 0.472 | 0.290 | 0.352 | 0.649 | 0.447 |

（a）Image （b）GT （c）Our_SCA （d）Contour （e）Our_SCAT

Figure 8. Visual results of failure case.

posed method. From Figure 8(e) and (b), we can see that the proposed method fails to produce the final saliency map. Figure 8(c) and (d) show the result of the direct learning and the intermediate process of twice learning. We can see that the error occurs in the process of the saliency contour detection, which means that the proposed method based on twice learning strategy heavily relies on the saliency contour detection. Even so, dividing the saliency detection process into two simpler process is reasonable and our ablation study and comparison results show the effectiveness of this method.

## 5. Conclusion

In this paper, we present a novel salient contour-aware based twice learning strategy for saliency detection. This method draws lessons from the process of information processing in human visual system and simulates the cognitive operation of humans in the state of painting copying. In this paper, the complex process of saliency detection is divided into saliency contour detection and saliency contour filling, and we use a simple and efficient network to achieve these two tasks. Ablation results prove that the proposed method based on the twice learning strategy outperforms direct learning. Comparison results show that our proposed method achieves competitive results. Moreover, we also propose a new saliency contour detection dataset. It will be published in our page[1].

[1]https://github.com/YanWei123/Salient-contour-aware-based-twice-learning-strategy-for-saliency-detection


## Acknowledgement

Thanks for National Engineering Laboratory for Video Technology - Shenzhen Division, Shenzhen Municipal Science and Technology Program (No. JCYJ20170818141146428), National Natural Science Foundation of China and Guangdong Province Scientific Research on Big Data (No.U1611461), Shenzhen Key Laboratory for Intelligent Multimedia and Virtual Reality (No.ZDSYS201703031405467).


## References


[1] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1597–1604, 2009. 6

[2] A. Borji, M. M. Cheng, H. Jiang, and J. Li. Salient object detection: A benchmark. *IEEE Transactions on Image Processing*, 24(12):5706–5722, 2015. 2

[3] A. Borji, D. N. Sihite, and L. Itti. *Salient Object Detection: A Benchmark*. Springer Berlin Heidelberg, 2012. 1

[4] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu. Global contrast based salient region detection. *IEEE TPAMI*, 37(3):569–582, 2015. 3, 4, 6

[5] Z. Deng, X. Hu, L. Zhu, X. xu, J. Qin, G. Han, and P.-A. Heng. R$\hat{3}$ net: Recurrent residual refinement network for saliency detection. 07 2018. 2, 6

[6] D.-P. Fan, M.-M. Cheng, Y. Liu, T. Li, and A. Borji. Structure-measure: A new way to evaluate foreground maps. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4548–4557, 2017. 6

[7] D. P. Fan, C. Gong, Y. Cao, B. Ren, and A. Borji. Enhanced-alignment measure for binary foreground map evaluation. 2018. 6

[8] L. Feng and M. Gleicher. Region enhanced scale-invariant saliency detection. In *IEEE International Conference on Multimedia Expo*, 2006. 3, 4

[9] S. Goferman, L. Zelnikmanor, and A. Tal. Context-aware saliency detection. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 34(10):1915, 2012. 6

[10] K. Huang, C. Zhu, and G. Li. Robust salient object detection via fusing foreground and background priors. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 2341–2345, Oct 2018. 6

[11] B. Jiang, L. Zhang, H. Lu, C. Yang, and M. H. Yang. Saliency detection via absorbing markov chain. In *IEEE International Conference on Computer Vision*, 2013. 2

[12] G. Lee, Y. W. Tai, and J. Kim. Deep saliency with encoded low level distance map and high level features. pages 660–668, 2016. 2

[13] G. Li and Y. Yu. Visual saliency based on multiscale deep features. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5455–5463, 2015. 3, 4, 6

[14] X. Li, F. Yang, H. Cheng, W. Liu, and D. Shen. Contour knowledge transfer for salient object detection. In *The European Conference on Computer Vision (ECCV)*, September 2018. 6

[15] X. Li, L. Zhao, L. Wei, M. H. Yang, F. Wu, Y. Zhuang, H. Ling, and J. Wang. Deepsaliency: Multi-task deep neural network model for salient object detection. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, 25(8):3919, 2015. 2

[16] Z. Luo, A. Mishra, A. Achkar, J. Eichel, and P. M. Jodoin. Non-local deep features for salient object detection. In *Computer Vision Pattern Recognition*, 2017. 6

[17] O. Ronneberger, P. Fischer, and T. Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Springer International Publishing, 2015. 6

[18] J. Shi, Q. Yan, L. Xu, and J. Jia. Hierarchical image saliency detection on extended cssd. *IEEE transactions on pattern analysis and machine intelligence*, 38(4):717–729, 2016. 3, 4, 6

[19] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *Computer Science*, 2014. 4

[20] L. Wang, H. Lu, Y. Wang, M. Feng, D. Wang, B. Yin, and X. Ruan. Learning to detect salient objects with image-level supervision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 136–145, 2017. 3, 4, 6

[21] C. Yang, L. Zhang, H. Lu, R. Xiang, and M. H. Yang. Saliency detection via graph-based manifold ranking. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3166–3173, 2013. 2, 3, 4, 6

[22] Q. Yao, H. Lu, Y. Xu, and W. He. Saliency detection via cellular automata. In *Computer Vision Pattern Recognition*, 2015. 6

[23] P. Zhang, D. Wang, H. Lu, H. Wang, and X. Ruan. Amulet: Aggregating multi-level convolutional features for salient object detection. In *ICCV*, Oct 2017. 2

[24] P. Zhang, D. Wang, H. Lu, H. Wang, and X. Ruan. Amulet: Aggregating multi-level convolutional features for salient object detection. In *ICCV 2017*, pages 202–211, 2017. 6

[25] P. Zhang, D. Wang, H. Lu, H. Wang, and B. Yin. Learning uncertain convolutional features for accurate saliency detection. In *ICCV*, Oct 2017. 6

[26] C. Zhu, X. Cai, K. Huang, T. H. Li, and G. Li. Pdnet: Prior-model guided depth-enhanced network for salient object detection. *arXiv preprint arXiv:1803.08636*, 2018. 6

[27] C. Zhu, K. Huang, and G. Li. An innovative saliency guided roi selection model for panoramic images compression. In *DCC*, pages 438–438, 2018. 1

[28] C. Zhu, G. Li, N. Li, X. Guo, W. Wang, R. Wang, C. Zhu, G. Li, N. Li, and X. Guo. An innovative saliency detection framework with an example of image montage. In *ACM MM*, 2017. 1

[29] C. Zhu, T. H. Li, and G. Li. Towards automatic wild animal detection in low quality camera-trap images using two-channeled perceiving residual pyramid networks. In *2017 ICCVW*, pages 2860–2864, Oct 2017. 2

[30] e. a. Zhu C, Huang K. Robust saliency detection via fusing center-enhanced foreground and background priors. 2019. 6